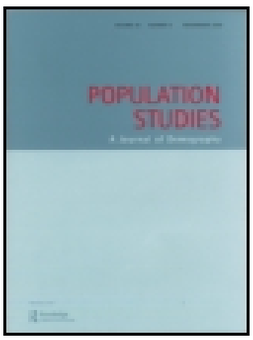
Using linked administrative and census data for migration research



Routledge
Taylor & Francis Group

Population Studies
A Journal of Demography

ISSN: 0032-4728 (Print) 1477-4747 (Online) Journal homepage: http://www.tandfonline.com/loi/rpst20

Using linked administrative and census data for migration research

Annemarie Ernsten, David McCollum, Zhiqiang Feng, Dawn Everington & Zengyi Huang

To cite this article: Annemarie Ernsten, David McCollum, Zhiqiang Feng, Dawn Everington & Zengyi Huang (2018): Using linked administrative and census data for migration research, Population Studies, DOI: 10.1080/00324728.2018.1502463

To link to this article: https://doi.org/10.1080/00324728.2018.1502463

Published online: 28 Aug 2018.

Population Studies, 2018
https://doi.org/10.1080/00324728.2018.1502463

Using linked administrative and census data for migration research

Annemarie Ernsten[1], David McCollum[1], Zhiqiang Feng[2], Dawn Everington[2] and Zengyi Huang[2]

[1]University of St Andrews, [2]The Longitudinal Studies Centre Scotland, University of Edinburgh

Migration is a core component of population change and is both a symptom and a cause of major economic and social phenomena. However, data limitations mean that gaps remain in our understanding of the patterns and processes of mobility. This is particularly the case for internal migration, which remains under-researched, despite being quantitatively much more significant than international migration. Using the Scottish Longitudinal Study, this paper evaluates the potential value of General Practitioner administrative health data from the National Health Service that can be linked into census-based longitudinal studies for advancing migration research. Issues relating to data quality are considered and, using the illustrative example of internal migration by country of birth, an argument is developed contending that such approaches can offer novel ways of comprehending internal migration, by shedding additional light on the nature of both movers and the moves that they make.

Keywords: administrative NHS GP health data; data linkage; internal migration; Scottish Longitudinal Study

[Submitted November 2017; Final version accepted June 2018]

Introduction

Migration is the key driver of population change at the local, regional, and national scales, but is also the hardest driver to measure and predict (Stillwell et al. 2011). This is especially true of internal migration, which despite being by far the quantitatively more significant phenomenon, has been subject to much less analytical scrutiny than international migration (Bell, Charles-Edwards, Kupiszewska, et al. 2015). This anomaly is noteworthy since the internal dynamics of population change are of considerable conceptual, policy, and commercial relevance (Smith et al. 2015). For example, around 5 per cent of the population of England and Wales (2.85 million people) moves between local authorities every year (Office for National Statistics (ONS) 2017a), and wide disparities in the incidence of internal migration exist at the global scale (Bell, Charles-Edwards, Ueffing, et al. 2015). As a spatially and socially selective process, internal migration is known to respond to economic change in specific ways, with an overall reduction in mobility during economic downturns (Saks and Wozniak 2011; Fielding 2012). However, much remains to be learned about how patterns and processes of internal migration have been shaped by the Great Recession of 2008 and ongoing economic uncertainty and austerity (Green and Shuttleworth 2015). Another topic that merits attention is the trajectory of internal migration in developed countries. Recent influential concepts in migration studies, such as the 'age of migration' (Castles et al. 2014) and the 'new mobilities' paradigm (Sheller and Urry 2006), have led to a common assumption that the current epoch is one of unprecedented mobility (Champion and Shuttleworth 2017a). However, this stands in contrast to an emerging body of evidence that points towards declining rates of internal migration, supposedly as a consequence of economic and demographic change, but also due to a wider societal shift towards so-called 'secular rootedness' (i.e., a more universal societal transition towards lower migration that cuts across multiple population subgroups) (Cooke 2011; Champion et al. 2017).

Efforts to understand these important issues better have long been hampered by a paucity of suitable data available to researchers concerning internal

mobility dynamics. Internal migration can be investigated through examination of population censuses, registers, and surveys, with each source having specific benefits and limitations (Bell, Charles-Edwards, Ueffing, et al. 2015). In the UK context, the best and thus most widely used data sources for research on internal migration have been the decennial censuses and the National Health Service Central Register (NHSCR), which is a health-based administrative register (Raymer et al. 2011). The NHSCR, which is based on General Practitioner (GP) registrations, is used by the official statistical offices in the UK to generate estimates of internal migration flows (ONS 2016). This source provides frequent and up-to-date information on moves. However, it contains a number of significant flaws: it undercounts some forms of mobility (such as short-distance moves and those made by young adults, especially men) and can shed light only on the origin, destination, age, and sex attributes of movers (Raymer et al. 2011). The decennial national census, on the other hand, contains a wealth of demographic information about movers, but is infrequent and picks up only individuals who have engaged in mobility sometime in the twelve months leading up to Census day (through the 'address one year ago' question). Nationally representative government surveys (such as the Labour Force Survey and Understanding Society in the UK's case) often contain time- and attribute-rich data that can be used to study migration. However, relatively limited sample sizes usually preclude detailed investigation of specific population subgroups or at subnational spatial scales (Stillwell et al. 2010).

Due to these limitations, researchers seeking to study internal migration have been faced with opting for either time-rich but attribute-poor administrative data, or attribute-rich but time-poor census data. However, recent linking of NHS GP registration data into the census-based Scottish Longitudinal Study (SLS) represents a potential new avenue for migration research, since it allows for study of the characteristics of both moves (via the NHSCR) and movers (through their linked census records). The novel methodological approach discussed in this paper thus allows, for the first time, a detailed analysis of the recent mobility patterns of a sizeable group of individuals at detailed geographies and over a significant period. This paper describes how this methodological approach operates and assesses the potential of such administrative data linked to census-based longitudinal studies to offer advances in migration studies. The following section offers a review of how internal migration is currently researched and considers how this can be enhanced using the innovative approach described in this paper.

Literature review

As discussed in the 'Introduction', an enhanced understanding of the dynamics of internal migration is necessary for a range of academic and practical reasons. The aim of this paper is not to contribute directly to recent research on patterns and processes of internal migration (Fielding 2012; Smith et al. 2015; Champion 2016; Champion et al. 2017), but rather to consider how new methodological innovations can build on this body of scholarship. As such, this section focuses mainly on *how* internal migration has been researched, rather than the results of these endeavours per se.

This paper seeks to demonstrate how the conventional difficulties faced by researchers drawing on either census-based longitudinal data sets or administrative data can be negated. Champion and Shuttleworth's recent studies of long-term trends in internal migration in England and Wales provide an apt illustration of these challenges. One of their studies uses the ONS Longitudinal Study (LS), the England and Wales sister study of the SLS, to examine changes in addresses over the ten-year periods between censuses from 1971 onwards (Champion and Shuttleworth 2017a). This provides a high level of detail on the characteristics of moves and movers but is limited by providing this type of information only at decadal intervals. Their other study (Champion and Shuttleworth 2017b) uses health administrative data to generate estimates of between-area moves, also from 1971 onwards. This approach benefits from annual as opposed to decennial information. However, it contains less information on the nature of moves and movers and omits mobility within Health Authority areas (i.e., shorter-distance mobility). The methodological perspective described in our paper seeks to overcome these limitations by allowing for identification of nearly all moves (via Scottish NHS GP registration data at unit postcode level) calculated on an annual basis, as well as the characteristics of movers (through the census-based SLS).

While this study represents one of only a few to link residential information from administrative health data to a census-based longitudinal data set explicitly, researchers have for some time explored the potential of combining data from multiple sources to understand internal migration better in

the absence of a population register in the UK. For example, Boden and Rees (2010) set out how various proxy measures of migration derived from administrative sources can enhance conventional ways of estimating immigration to local areas in England, resulting in ONS adopting these methods in their estimates of local immigration flows. Furthermore, Raymer et al. (2011) outline a method for combining NHSCR data with census data to generate a synthetic database of intercensal migration patterns. This approach involves supplementing information from the NHSCR with more detailed information from the censuses using log-linear modelling procedures. A well-established, important limitation of using administrative health data to assess migration is the age–sex bias in the propensity to register with a doctor after moving (Ogilvy 1980; Boden et al. 1992). Young adults, especially males, are less likely to reregister after moving and may take longer to do so than women and older people. Raymer et al. (2011) propose weighting procedures to overcome this challenge, whereas others actually use these techniques to correct undercounts in the migration rates of young adult males (ODPM 2002). The potential of weighting to improve the veracity of migration statistics is an issue considered in this analysis. A valuable extension of Raymer et al.'s (2011) method—offered by the approach described in this paper—is that longitudinal census data are used, allowing for individuals to be followed over time and thus for consideration of the roles of age, period, and cohort effects in migration trends (Findlay et al. 2015). Another interesting avenue offered by the approach discussed here is that the health administrative data in Scotland are available at finer geographies than is the case elsewhere in the UK, thus enabling short-distance mobility to be incorporated into the analysis. As such, all changes in address can be assessed as opposed to just those which span higher-level administrative boundaries.

A key issue when combining census-based and administrative data is the extent to which the information on the locations and migration experiences of individuals provided by each source correspond. The effectiveness of this 'matching' process is an issue considered at length in this analysis. A precedent can be gleaned from existing research in this respect, as NHSCR data have been linked into the ONS LS in the past. Smallwood and Lynch (2010) test the extent to which the census and NHSCR provide consistent information on individual's Health Authority of residence using the 'snapshot' of Census day 2001 (Sunday 29 April). They conclude that the census-based longitudinal study and health administrative data are very good 'matches' in terms of individuals' locations, with 96 per cent of ONS LS members recorded as being in the same Health Authority area as stated in their NHSCR record (although the figure was lower for young men and students). However, this is a one-off study, as NHSCR-derived records are not routinely uploaded into the ONS LS. Another caveat here is that, as mentioned earlier, the available England and Wales health administrative data only include moves between Health Authority areas, whereas the newly available Scottish data include all unit postcode changes (postcode units are unique references and identify an average of 15 addresses). As such, the degree of matching at health area level will be considerably higher than that between unit postcode levels, given the higher-level geographies involved in the former (e.g., the average population of local health areas in England is 350,000; National Audit Office 2012).

Drawing on the Health Card Registration System (Northern Ireland's equivalent of the NHSCR) and the Northern Ireland Longitudinal Study, Barr and Shuttleworth (2012) also consider the issue of the quality of matching between census and health administrative data (at the Super Output Area level). Their approach involves focusing on movers within Northern Ireland and modelling the likelihood of: (a) an exact location match in their census and health administrative records; (b) a lagged match, whereby it takes over a year for movers' health administrative data location to match their census location; and (c) no locational match being found. The study, which excludes students, finds that the residential movements of men, those in good health, and residents of urban and relatively deprived areas are most likely to be missed using health administrative data, via either late or non-reporting of address changes. The research discussed in our paper explores these issues in the Scottish context. Critically, our investigation also details how weighting procedures can overcome the limitations of late and non-response in health administrative data for migration research. The following section sets out the methodological approach used in our analysis.

Data

The core focus of this research is the matching of information from the SLS with mobility-related information held by the NHSCR, and consideration of the utility of this process for migration research.

In this section we provide a brief description of the core data sets used in this approach and how they are matched to each other. The following section considers the quality of this process in terms of the extent to which certain moves and movers may be under-represented in the data. Next, we examine the ability of weighting procedures to address these issues. Finally, we use the example of internal migration rates in Scotland by country of birth to illustrate the usefulness of this type of approach in migration studies.

The SLS is a large-scale linkage study based on information from Scottish censuses from 1991 onwards. The study is based on 20 semi-random birthdates. Four of these match the birthdates in the England and Wales LS and the remaining 16 are chosen randomly from the remaining 361 days in the year (362 in a leap year), using probabilities derived from the daily distribution of births. Following the removal of dummies and duplicate records, about 5.3 per cent of the Scottish population is covered in the sample, equating to around 265–270,000 members (Hattersley et al. 2007; Boyle et al. 2009). Data are collected on SLS members over time and their records are continuously updated through the linkage of vital event registration and NHSCR data (Hattersley and Boyle 2007). The linking together of records from various administrative sources over time is an integrated part of the SLS. The record linkage exercise is carried out using the NHSCR database of all residents in Scotland who have registered with an NHS GP. This is the most inclusive 'register' of the population in Scotland. Names and dates of birth are two of the basic pieces of information required by the NHSCR to allow an individual to be identified in the database and then for SLS members to be 'flagged' so that they can be linked to the SLS. Linking to non-census data sets and linking between censuses both depend on the tracing of SLS members in the NHSCR. The tracing of SLS members in the NHSCR is carried out using a combination of exact matching, probability matching, and manual matching (Hattersley and Boyle 2008).

The SLS only recently (in 2016) received permission to add historical NHSCR GP postcode data (starting from 1 January 2000) into the SLS. This recent development means that research on internal migration in Scotland can now be carried out using health administrative data linked to census-based longitudinal studies, akin to that conducted in Northern Ireland (Barr and Shuttleworth 2012). However, the Scottish data have the important additional advantage of enabling analysis of short-distance moves. The way that health administrative data are incorporated into the census-based longitudinal studies of the other parts of the UK results in only moves spanning Health Authority boundaries or Super Output Areas being recorded (see ONS 2017b for an illustration of UK statistical geographies). In Scotland, information at unit postcode level from the Community Health Index (CHI) system is now fed into the NHSCR. Postcode-level data cannot be directly assessed by researchers due to the risk of statistical disclosure, but SLS support staff can derive variables for moves (such as distance of move) for researchers to use; this allows for analysis of short-distance moves without the risk of disclosure. As such, the recent linking of these data into the SLS now enables analysis of moves at postcode unit level upwards, as opposed to merely the longer-distance moves that cross health administrative boundaries. It should be noted that the analysis described in this paper is based on a test version of NHSCR GP postcode data, which has subsequently been revised. While this affects only a small proportion of the data, the data set now available to researchers is slightly different. This, and the specific sample definition and methods used in our study, means that the results from future analyses may not exactly correspond with those described here.

Effectiveness of the matching process

The previous section briefly described the SLS and the health administrative data that constitute the basis of this new resource for migration research. An issue of critical importance in this respect relates to the effectiveness of the matching between the two data sets. Ideally everyone should be detected as being at the same postcode on the Census days (Sunday 29 April 2001 and Sunday 27 March 2011) according to their census enumeration and NHS GP data. This is the case for 85 per cent of the SLS members included in this analysis (those of working age, 16–64, at the 2011 Census; $N =$ 174,258). The fact that this is lower than the 96 per cent found by Smallwood and Lynch (2010) using the ONS LS is most likely due to the finer geographies used in our analysis (postcodes vs. Health Authority boundaries).

Of those with non-matching census and health administrative records, a relatively small number of individuals (1.1 per cent) do not have NHS GP data, mainly because these SLS members are not traced at the NHSCR. The intercensal mobility of these individuals cannot therefore be detected.

Table 1 Comparison of the postcodes recorded by the 2011 Census vs. the NHSCR, Scotland

Category	Locational information: SLS vs. NHSCR at Census day	Percentage
Exact match	Same postcode location at Census day	85.0
Delayed match: NHSCR postcode matches 2011 Census postcode	Within 6 months	0.6
	Within 7–12 months	0.9
	Within 13–24 months	1.3
	Within 25–36 months	1.7
	Over 36 months	0.1
No match	In SLS and traced in NHSCR but never a postcode match	9.3
	In SLS but not traced in NHSCR	1.1
Total		100.0

Note: Sample consists of SLS study members aged 16–64 on Census day 2011 ($N = 174{,}258$).
Source: SLS.

Another, more significant group for whom researching mobility behaviour is complex consists of those who display a delayed match between their census record and their location according to the NHSCR. In total, 13.9 per cent of the sample used in this research are present in both data sets but do not have an exact unit postcode match on Census day. For the majority of these (9.3 per cent of the overall sample) the census enumeration postcode is not found in the NHS GP postcode history. For the remainder (4.5 per cent of the overall sample), the census enumeration postcode is found at a later date in the NHS GP postcode history, almost always within three years (Table 1). For the purposes of the illustrative example using country of birth data in this paper, the effective study sample is the 85.0 per cent of the SLS members of working age (16–64 years) at the 2011 Census with exact matches between the 2011 Census and NHS GP postcodes, plus the 0.6 per cent with a delayed match within six months of Census day.

Suitability for migration research, under-reporting bias, and weighting

Having deduced that the matching rate is high at fine geographic scales, it is reasonable to assert that the linking of administrative health data to census-based longitudinal data sets represents an avenue of potential advancement in internal migration research. However, as migration is inherently a socially and spatially selective process, it is important to consider which movers and types of moves are under-represented by this approach. This issue, and how it can be addressed through weighting procedures, is the focus of this section.

Figure 1 shows matching rates (up to six months after Census day 2011) by age and sex, and confirms that matching rates are lowest for those aged in their 20s (especially males), as people in this age group are less likely to register with a new doctor after moving, or take longer to do so (Smallwood and Lynch 2010; Raymer et al. 2011).

While Figure 1 provides a useful illustration that some moves are likely to be under-represented in health administrative data, inferential statistics can give a fuller account of the factors that are likely to result in an individual's mobility being detected in the approach discussed in this paper. Table 2 presents the results of a binary logistic model, which examines the likelihood of an individual's NHSCR GP postcode information matching their census postcode (up to six months after Census day). The results confirm expectations regarding the significant impacts of age and sex on the likelihood of effective address matching between administrative health and census data. Women's matching rates are much higher than men's, and the mobility of those aged in their 20s is least likely to be detected using this approach. Additionally, unpartnered individuals are less likely to be matched than those who are partnered, and students display low matching rates relative to those with other employment statuses, especially homemakers and the long-term sick and disabled. In line with Barr and Shuttleworth (2012), geography matters in that residents of less deprived areas are in general more likely to be matched than those of more deprived ones, although interestingly the top and bottom quintiles are not statistically different in this respect. Finally, matching rates are relatively low in large urban areas, potentially because residents can change address without necessarily needing to change doctor (at least initially). Two exceptions to the higher matching rates outside large

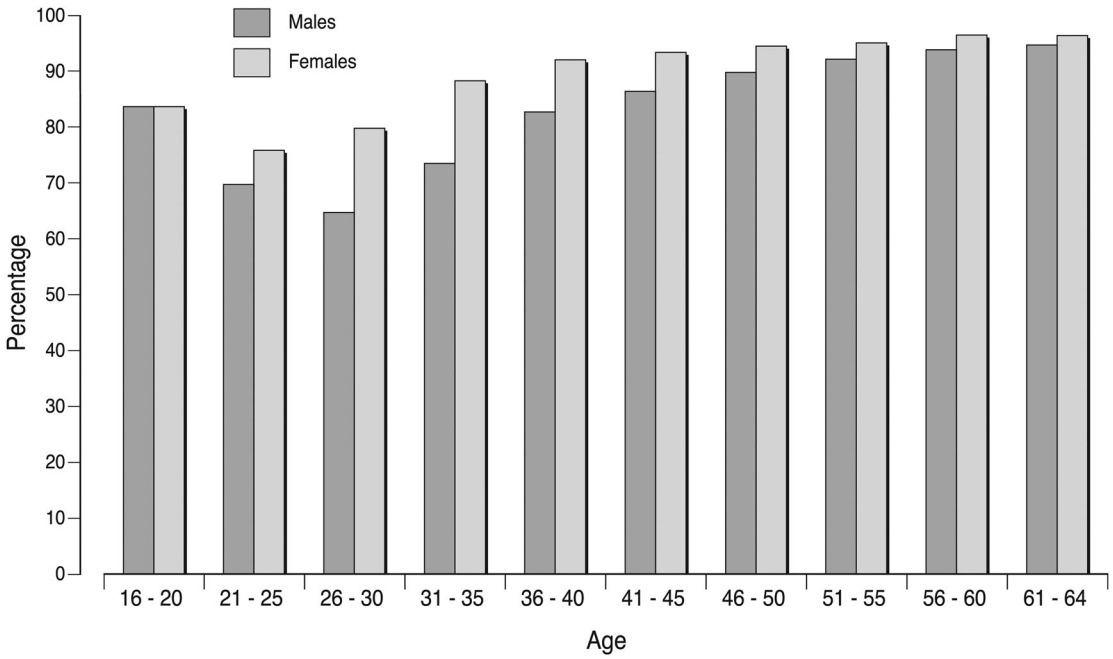


Figure 1 SLS and NHSCR postcode matching rates (up to six months after 2011 Census day) by age and sex, Scotland

Note: Sample consists of SLS study members aged 16–64 on Census day 2011 ($N = 174{,}258$).
Source: SLS.

urban areas are very remote small towns and very remote rural areas, possibly because residents of these locations use health services less frequently than their less geographically remote counterparts (Hine and Kamruzzaman 2012).

The discussion thus far has established that linking administrative health data to census-based longitudinal data sets represents a potentially valuable avenue for migration research, but that such approaches suffer from a systematic under-reporting of some forms of mobility. As suggested by Raymer et al. (2011), weighting procedures represent a potential means of overcoming this limitation. Our analysis creates weights to demonstrate how the bias created by the under-reporting of moves by particular groups can be addressed.

Weighting and data imputation are both possible responses to missing data. Weighting is normally applied in the case of unit non-response (complete absence of a record for an individual at a specific time), whereas imputation usually occurs in instances of item non-response, that is, partially missing data pertaining to an individual (Lynn 1996). In this case, weighting rather than imputation is used to adjust for those people who are missing from the matched postcode subgroup, as we know that such people are not missing at random. Having up-to-date GP registration information will depend on several different characteristics. As evident in Figure 1 and Table 2, the probability of postcode matching between census and health administrative data is highly age and sex dependent. As such, weights are created to account for these factors using an inverse probability method, which multiplies units by the distribution of their age- and sex-defined subgroup in the total sample population, divided by the distribution in the sample with matched postcode data. For example, if the subgroup accounted for one-fifth of the sample with matched postcode data and one-quarter of the total sample population, then the corresponding weight would be $5 \div 4 = 1.25$, thus increasing the presence of this group. A further factor is applied so that the total sample size of the matched postcode subsample after weighting is the same as the size of the total sample. The weights produced by this process are displayed in Table 3 and this type of approach could be of value to other researchers using the SLS for similar purposes. As can be seen, those aged in their 20s have higher weights (increasing the representation of these groups), whereas older age groups have values closer to one. For completeness, weights are calculated for both the 2001 and 2011 Census years (note the better matching rates, indicated by lower weights, in the latter) and scaled weights are included to account for variations in sample sizes between the two censuses. The 'normal' weights refer to all sample members that were present in the 2001 or

Table 2 Binary logistic model: predictors of address match between census and health administrative data (up to six months after Census day 2011), Scotland

Variable		Odds ratio	Standard error
Sex	Male (ref)	1.000	
	Female	1.848***	0.05
Age group	21–30 (ref)	1.000	
	16–20	2.891***	0.11
	31–40	1.569***	0.04
	41–50	3.048***	0.09
	51–64	5.305***	0.17
Partnered	Yes (ref)	1.000	
	No	0.698***	0.01
Employment status	Full-time employed (ref)	1.000	
	Part-time employed	1.330***	0.03
	Unemployed	0.994	0.03
	Student	0.837***	0.03
	Homemaker	1.689***	0.09
	Long-term sick or disabled	1.521***	0.06
	Other (including retired)	1.000	0.04
Scottish Government Urban Rural Classification[1]	Large urban areas (ref)	1.000	
	Other urban areas	1.455***	0.03
	Accessible towns	1.585***	0.05
	Remote small towns	1.878***	0.12
	Very remote small towns	1.131	0.08
	Accessible rural areas	1.150***	0.03
	Remote rural areas	1.184***	0.06
	Very remote rural areas	1.014	0.05
Scottish Index of Multiple Deprivation	Least deprived quintile (ref)	1.000	
	4th quintile	0.864***	0.02
	Middle quintile	0.864***	0.02
	2nd quintile	0.927**	0.02
	Most deprived quintile	0.958	0.02
	Constant	2.428***	0.07

[1]Areas classified according to population size (thousands) and distance in time to nearest large settlement: Large urban areas (>125), Other urban areas (10–125), Accessible towns (3–10, <30 minutes), Remote small towns (3–10, 30–60 minutes), Very remote small towns (3–10, 60 + minutes), Accessible rural areas (<3, <30 minutes), Remote rural areas (<3, 30–60 minutes), Very remote rural areas (<3, 60+ minutes).
Notes: Ref is the reference category. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.
Source: Authors' analysis of SLS 2011 ($N = 172{,}721$).

Table 3 Illustrative example of age and sex weights used to account for missing data in the health administrative data linked to the 2001 and 2011 Censuses, Scotland

	Normal weights		Scaled weights	
	2001	2011	2001	2011
Males aged				
16–20	1.290	1.191	1.303	1.178
21–30	1.595	1.469	1.612	1.454
31–40	1.376	1.258	1.390	1.245
41–50	1.227	1.128	1.240	1.116
51–64	1.164	1.072	1.176	1.061
Females aged				
16–20	1.287	1.172	1.300	1.160
21–30	1.359	1.236	1.373	1.223
31–40	1.181	1.092	1.194	1.081
41–50	1.137	1.060	1.149	1.049
51–64	1.122	1.046	1.133	1.035

Source: Authors' analysis of SLS data.

2011 Censuses, whereas the 'scaled' weights are for analyses involving only SLS members that were present in both censuses.

Suitability for migration research: an illustrative example

Having considered the potential suitability of weighted administrative health data linked to a census-based longitudinal study for migration research, the discussion now briefly turns to an illustrative example of its application: a time series of internal migration trends in Scotland. As touched on earlier, much uncertainty remains with regard to how internal migration in advanced economies has responded to recent economic change (Green and Shuttleworth 2015), as well as to longer-term societal shifts regarding mobility (Champion et al. 2017). The approach discussed in this paper can help to shed light on this issue in the Scottish context. Another pertinent issue (see Figure 2) relates to the under-researched internal mobility experiences of international migrants after arriving in their host country (Catney and Finney 2012; Jivraj et al. 2012). Internal mobility patterns of international migrants may be dissimilar to those of long-term residents with similar characteristics, since the mobility trends of the former group may represent part of a longer period of adjustment arising from their initial entry into the country (Trevena et al. 2013). The sample in this study is SLS members aged 16–64 at the 2011 Census, whose records are traced by NHSCR and whose location according to the census and administrative health data matches on Census day or within six months of it. This equates to 151,498 individuals in 2011, with fewer before and after this date (126,755 in 2001; 138,089 in 2015) as people enter and leave the survey through birth, deaths, and migration to and from Scotland. The migration rate of this sample is measured annually from 2001 to 2015 and is defined as the number of moves in a given year divided by the study sample population of that year.

In line with the experience of many other countries (Champion et al. 2017), the trends in Figure 2 suggest that Scots, who represent over four-fifths of the population of Scotland, appear to be becoming gradually less mobile. However, international migrants in Scotland seem to show distinct trends. The internal migration rates of European Union nationals, in particular, exceed those of other groups during and following the 2008 recession. Although not presented here, by allowing for analysis of short-distance moves, the research also uncovers interesting differences in residential mobilities between immigrant

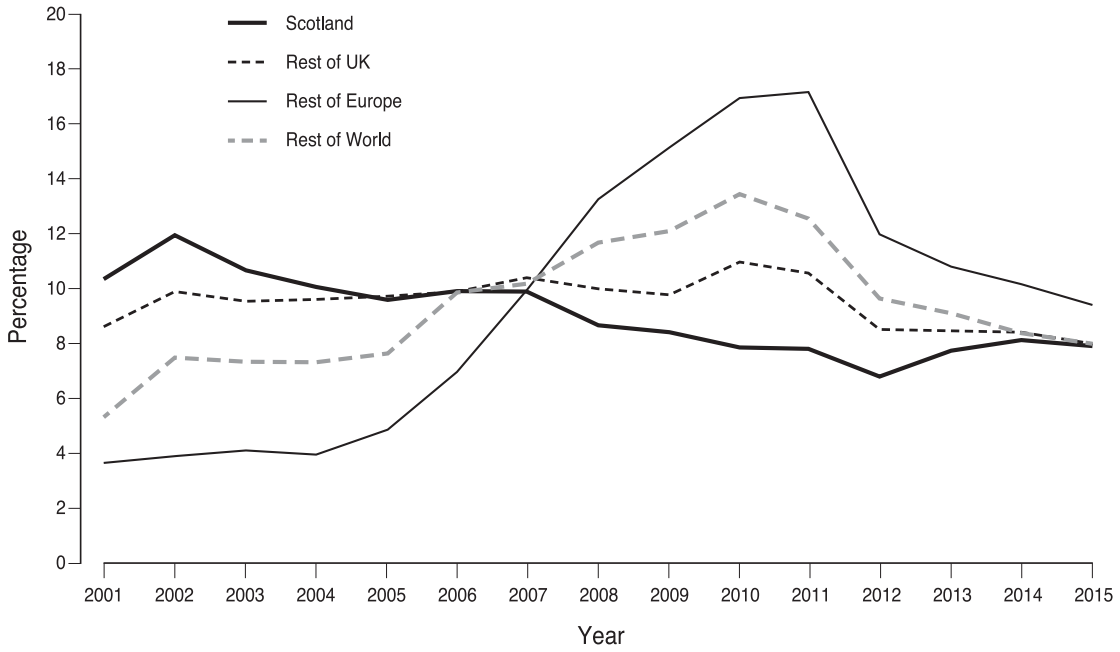


Figure 2 Internal migration rates in Scotland, 2001–15, by country of birth

Notes: Sample consists of SLS members of working age (16–64) in 2011, where 2011 Census postcode matches NHS GP postcode on Census day or up to six months following ($N = 151,498$). Note the smaller sample in Figure 2 than Figure 1, which is a consequence of country of birth information being missing for some sample members.

Source: Authors' analysis of SLS.

groups. These preliminary findings are ripe for further, more detailed analysis. While not the core focus of this paper, these trends provide an apt illustration of the intriguing research themes that the novel methodological approach described here permits.

Conclusions

As detailed by the analysis described in this paper, the recent (2016) linking of health administrative data into the SLS presents a valuable opportunity to examine patterns and processes of internal migration. While this innovative approach clearly offers a major advancement in how migration can be researched in the Scottish context, it also holds lessons for wider demographic research. Most significantly, it demonstrates the value of data linkage to aid understanding of migratory patterns and processes (ONS 2009). Combined with ongoing advances in longitudinal data sets and analytical procedures (Findlay et al. 2015), this investigation provides support for efforts to encourage enhanced linking of administrative data into major longitudinal studies. Such moves can facilitate the development of new research agendas. In the field of migration studies, for example, administrative data linked to census-based longitudinal studies can aid research into such pertinent issues as the spatial–social mobility nexus (Favell and Recchi 2011), the relationship between internal migration and business cycles (Saks and Wozniak 2011), the internal migration of international migrants (Catney and Finney 2012), and the question of whether there is indeed a fundamental shift towards 'secular rootedness' (Cooke 2011) in advanced economies.

Finally, it is worth emphasizing that the focus on the suitability of administrative data for assessing migration is likely to become ever more prominent as national statistics authorities begin to move away from traditional censuses. In the UK, for example, the next census (in 2021) will be predominantly online and will make increased use of administrative data and surveys both to enhance the statistics from the 2021 Census and improve statistics between censuses (ONS 2017c). Given the extensive use of the conventional decennial national census to research migration, the threat of its demise in the future presents some significant challenges to the research community, not least because the approach discussed in this paper would not be possible without census-based longitudinal studies. Going forward, much rests on the quality of administrative data that can inform migration research. In a best-case scenario, this could be so rich that it even negates the need for census-based data. However, a further challenge relates to the availability of such data. In England and Wales, for example, health administrative data would theoretically allow for the same type of analysis as described here, but the data are unfortunately unavailable to researchers at the unit postcode scale. In the context of questions over its richness and accessibility, and in the absence of a population register, the utility of administrative data for research in the social sciences is likely to come under increasing scrutiny. These discussions will inevitably involve legal and ethical conundrums associated with data sharing, alongside practical questions relating to how researchers access and use these data-rich resources. By shedding light on how the linkage of administrative health data into a census-based longitudinal study can aid migration research, we hope that this paper contributes to these timely and significant debates.

Notes and acknowledgements

1 Please direct all correspondence to David McCollum, Geography & Sustainable Development, Irvine Building, University of St Andrews, North Street, St Andrews KY16 9AL, Fife, Scotland, UK; or by E-mail: David.McCollum@st-andrews.ac.uk

2 This research was funded through the Economic and Social Research Council (ESRC) Secondary Data Analysis Initiative (Grant number ES/N011430/1).

3 We acknowledge the help provided by the staff of the Longitudinal Studies Centre—Scotland (LSCS). The LSCS is supported by the ESRC/JISC, the Scottish Funding Council, the Chief Scientist's Office, and the Scottish Government. We are responsible for the interpretation of the data. Census output is Crown copyright and is reproduced with the permission of the Controller of HMSO and the Queen's Printer for Scotland.

References

Barr, Paul and Ian Shuttleworth. 2012. Reporting address changes by migrants: The accuracy and timeliness of reports via health card registers, *Health and Place* 18 (3): 595–604.

Bell, Martin, Elin Charles-Edwards, Dorota Kupiszewska, Marek Kupiszewski, John Stillwell, and Yu Zhu. 2015. Internal migration data around the world: Assessing contemporary practice, *Population, Space and Place* 21 (1): 1–17.

Bell, Martin, Elin Charles-Edwards, Philipp Ueffing, John Stillwell, Marek Kupiszewski, and Dorota Kupiszewska. 2015. Internal migration and development: Comparing migration intensities around the world, *Population and Development Review* 41(1): 33–58.

Boden, Peter and Philip Rees. 2010. Using administrative data to improve the estimation of immigration to local areas in England, *Journal of the Royal Statistical Society: Series A Statistics in Society* 173(4): 707–731.

Boden, Peter, John Stillwell, and Philip Rees. 1992. How good are NHSCR data? in J. Stillwell, P. Rees, and P. Boden (eds), *Migration Processes and Patterns. Volume 2: Population Redistribution in the United Kingdom*. London: Belhaven Press, pp. 13–27.

Boyle, Paul, Peteke Feijten, Zhiqiang Feng, Lin Hattersley, Zengyi Huang, Joan Nolan, and Gillian Raab. 2009. Cohort profile: the Scottish Longitudinal Study (SLS), *International Journal of Epidemiology* 38(2): 385–392.

Castles, Stephen, Hein de Haas, and Mark Miller. 2014. *The Age of Migration: International Population Movements in the Modern World* (5th edition). Basingstoke: Palgrave MacMillan.

Catney, Gemma and Nissa Finney. 2012. Minority internal migration in Europe, in N. Finney and G. Catney (eds), *Minority Internal Migration in Europe*. Farnham: Ashgate, pp. 1–12.

Champion, Tony. 2016. Internal migration and the spatial distribution of population, in T. Champion and J. Falkingham (eds), *Population Change in the United Kingdom*. London: Rowman and Littlefield, pp. 125–142.

Champion, Tony and Ian Shuttleworth. 2017a. Are people changing address less? An analysis of migration within England and Wales, 1971–2011, by distance of move, *Population, Space and Place* 23(3): 1–13.

Champion, Tony and Ian Shuttleworth. 2017b. Is longer-distance migration slowing? An analysis of the annual record for England and Wales since the 1970s, *Population, Space and Place* 23(3): 14–28.

Champion, Tony, Thomas Cooke, and Ian Shuttleworth (eds). 2017. *Internal Migration in the Developed World: Are we Becoming Less Mobile?* Abingdon: Routledge.

Cooke, Thomas. 2011. It is not just the economy: Declining migration and the rise of secular rootedness, *Population, Space and Place* 17(3): 193–203.

Favell, Adrian and Ettore Recchi. 2011. Social mobility and spatial mobility, in A. Favell and V. Guiraudon (eds), *Sociology of the European Union*. Houndmills: Palgrave Macmillan, pp. 50–75.

Fielding, Tony. 2012. *Migration in Britain: Paradoxes of the Present, Prospects for the Future*. Cheltenham: Edward Elgar.

Findlay, Allan, David McCollum, Rory Coulter and Vernon Gayle. 2015. New mobilities across the life course: A framework for analysing demographically linked drivers of migration, *Population, Space and Place* 21(4): 390–402.

Green, Anne and Ian Shuttleworth. 2015. Labour markets and internal migration, in D. Smith, N. Finney, K. Halfacree, and N. Walford (eds), *Internal Migration: Geographical Perspectives and Processes*. Farnham: Ashgate, pp. 65–80.

Hattersley, Lin and Paul Boyle. 2007. *The Scottish Longitudinal Study: An Introduction*. SLS Technical Working Paper 1. Edinburgh/St Andrews: Longitudinal Studies Centre Scotland.

Hattersley, Lin and Paul Boyle. 2008. *The Scottish Longitudinal Study, a Technical Guide to the Creation, Quality and Linkage of the 2001 Census SLS Sample*. Edinburgh/St Andrews: Longitudinal Study Centre Scotland & General Register Office for Scotland.

Hattersley, Lin, Gillian Raab, and Paul Boyle. 2007. *The Scottish Longitudinal Study, Tracing Rates and Sample Quality for the 1991 Census SLS Sample*. Edinburgh/St. Andrews: Longitudinal Studies Centre Scotland and General Register Office for Scotland.

Hine, Julian and Md Kamruzzaman. 2012. Journeys to health services in Great Britain: an analysis of changing travel patterns 1985–2006, *Health and Place* 18(2): 274–285.

Jivraj, Stephen, Ludi Simpson, and Naomi Marquis. 2012. Local distribution and subsequent mobility of immigrants measured from the School Census in England, *Environment and Planning A* 44(2): 491–505.

Lynn, Peter. 1996. Weighting for non-response, in R. Banks, J. Fairgrieve, L. Gerrard, T. Orchard, C. Payne, and A. Westlake (eds), *Survey and Statistical Computing*. Proceedings of the second ASC international conference, Imperial College, London, UK, September 11–13 1996. pp. 205–214.

National Audit Office. 2012. *Healthcare Across the UK: A Comparison of the NHS in England, Scotland, Wales and Northern Ireland*. Report by the Comptroller and Auditor General.

ODPM. 2002. *Development of a Migration Model*. Report prepared by the University of Newcastle upon Tyne, the University of Leeds, and the Greater London Authority/London Research Centre. ODPM. London. Available: https://www.researchgate.net/profile/Stamatis_Kalogirou/publication/259308761_Development_of_a_Migration_Model_Analytical_and_Practical_Enhancements/links/00b4952af09f4101aa000000/Development-of-a-Migration-Model-Analytical-and-Practical-Enhancements.pdf (accessed: 8 February 2018).

Office for National Statistics. 2009. *A Summary of Administrative Data Sources and their Potential to*

Inform Statistics on Migration and Population. Available: file:///C:/Users/dm82/Downloads/statementofadminsources_tcm77-185902%20(1).pdf (accessed: 8 September 2017).

Office for National Statistics. 2016. *NHSCR Data: Quality Assurance of Administrative Data used in Population Statistics*. Dec 2016. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/methodologies/nhscrdataqualityassuranceofadministrativedatausedinpopulationstaticsdec2016 (accessed: 28 July 2017).

Office for National Statistics. 2017a. Population Estimates for UK, England and Wales, Scotland and Northern Ireland: mid-2016. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/latest (accessed: 28 July 2017).

Office for National Statistics. 2017b. *Historical Representation of UK Statistical Geographies*. Available: http://geoportal.statistics.gov.uk/datasets?q=Hierarchical%20Representation%20of%20UK%20Statistical%20Geographies%20(September%202017)&sort=name (accessed: 19 October 2017).

Office for National Statistics. 2017c. Annual assessment of ONS's progress towards an Administrative Data Census post-2021 Available: file:///C:/Users/dm82/Downloads/annualassessment2017v04.pdf (accessed: 8 September 2017).

Ogilvy, Audrey. 1980. Inter-regional migration since 1971: An appraisal of data from the National Health Service Central Register and Labour Force Surveys. Office of Population Censuses and Surveys, Occasional Paper 16.

Raymer, James, Peter Smith, and Corrado Giulietti. 2011. Combining census and registration data to analyse ethnic migration patterns in England from 1991 to 2007, *Population, Space and Place* 17(1): 73–88.

Saks, Raven and Abigail Wozniak. 2011. Labor reallocation over the business cycle: New evidence from internal migration, *Journal of Labor Economics* 29(4): 697–739.

Sheller, Mimi and John Urry. 2006. The new mobilities paradigm, *Environment and Planning A* 38(2): 207–226.

Smallwood, Steve and Kevin Lynch. 2010. An analysis of patient register data in the longitudinal study-what does it tell us about the quality of the data? *Population Trends* 141: 151–169.

Smith, Darren, Nissa Finney, Keith Halfacree, and Nigel Walford. 2015. Introduction, in D. Smith, N. Finney, K. Halfacree, and N. Walford (eds), *Internal Migration: Geographical Perspectives and Processes*. Farnham: Ashgate, pp. 1–14.

Stillwell, John, Peter Boden, and Adam Dennett. 2011. Monitoring who moves where: Information systems for internal and international migration, in J. Stillwell and M. Clarke (eds), *Population Dynamics and Projection Methods*. London: Springer, pp. 115–140.

Stillwell, John, Adam Dennett, and Oliver Duke-Williams. 2010. Interaction data: Definitions, concepts and sources, in J. Stillwell, A. Dennett, and O. Duke-Williams (eds), *Technologies for Migration and Commuting Analysis: Spatial Interaction Data Applications*. Hershey: IGI Global, pp. 1–30.

Trevena, Paulina. Derek McGhee, and Sue Heath. 2013. Location, location? A critical examination of patterns and determinants of internal mobility among post-accession Polish migrants in the UK, *Population, Space and Place* 19(6): 671–687.